In one such embodiment, therefore, an error metric S^2, defined as follows:

$$S^2 = \frac{\sum_{i>j}^{n} \left(|x_i - x_j| - d_{ij} \right)^2}{\sum_{i>j}^{n} (d_{ij})^2}$$

Please replace the paragraph beginning at **page 5, line 3** with the following rewritten paragraph:

It may in some cases be advantageous to use alternative versions of S^2. For example, there are alternatives for this formula from the distance geometry literature which are also suitable for use in conjunction with the invention. Several of these can be found in "The Theory and Practice of Distance Geometry", T.F. Havel, I.D. Kuntz, and G.M. Crippen, Bull. Math. Biol., vol. 45, pp. 665-720 (1983), the entire disclosure of which is hereby incorporated by reference in its entirety. One such alternative function is:

$$S^2 = \sum_{i>j}^{n} \left[1 - \left(\frac{x_i - x_j}{d_{ij}} \right)^2 \right]^2$$

Please replace the paragraph beginning at **page 5, line 13** with the following rewritten paragraph:

Havel et al point out that this function exhibits good behavior for optimization purposes. Another possible function is:

$$S^2 = \sum_{i>j}^{n} \left\{ 1 - \left[\max(x_{ij} / d_{ij}, d_{ij} / x_{ij}) \right]^2 \right\}^2 .$$

Please replace the paragraph beginning at **page 16, line 30** with the following rewritten paragraph:

Once the maximum overlap value is determined, a molecular similarity score Sim_{AB} can be defined on the interval from 0 to 1 by normalizing the maximum overlap measured as follows:

$$Sim_{AB} = \frac{S_{\max,AB}}{\sqrt{S_{\max,AA}\, S_{\max,BB}}}$$

Please replace the paragraph beginning at **page 22, line 30** with the following rewritten paragraph:

To perform this comparison, string A and string B are oriented with their centers aligned. Then, the position of string B is shifted to align, as closely as possible, common atom pairs between the two strings. The amount of this shift Δx_B is calculated as follows:

$$\Delta x_B = \frac{1}{N_{match}} \sum_{\substack{Common \\ atom \\ pairs}} (x_A - x_B)$$

Please replace the paragraph beginning at **page 23, line 7** with the following rewritten paragraph:

After aligning the strings in this way, the squares of the linear offsets between all atom pairs of the same class in string A and string B is computed to produce a sum-squared-deviation (SSD) as follows:

$$SSD = \sum_{\substack{Common \\ atom \\ pairs}} (x_A - x_B)^2$$

Please replace the paragraph beginning at **page 23, line 21** with the following rewritten paragraph:

Figures 11 and 12 illustrate the results of a comparison between a first compound, denoted compound A with two other compounds, denoted B1 and B2. Similarity calculations were performed using both 3D atomic coordinates to derive 1D representations, and also using 2D topological information to derive 1D representations. Figure 11 shows the result of the comparison between compound A and compound B1 when 3D and 2D information was used as a starting point. Figure 12 shows the result of the comparison between compound A and compound B2 when 3D and 2D information was used as a starting point. Although graphs of overlap as a function of offset are shown in Figures 11 and 12 for illustrative purposes, it will be appreciated that in accordance with the above described techniques, most of the computations needed to generate such graphs are not required to be performed in order to produce the desired similarity measure. Using the equation:

$$Sim_{AB} = \frac{S_{\max,AB}}{\sqrt{S_{\max,AA}\, S_{\max,BB}}}$$

the similarity value Sim_{AB1} for compounds A and B1 is 0.564, when 3D coordinates are used to derive the 1D representations, and is 0.529 when 2D topology is used to derive the 1D representations. In addition, the similarity value Sim_{AB2} for compounds A and B2 is 0.709, when 3D coordinates are used to derive the 1D representations, and is 0.775 when 2D topology is used to derive the 1D representations.

Please replace the paragraph beginning at **page 25, line 23** with the following rewritten paragraph:

Although this procedure is fast, one problem with it is the fact that there are usually a large number of bin aligned orientations to consider. This number can be reduced in a manner analogous to that described above by computing upper bounds for each bin aligned position, and then eliminating from consideration those bin aligned orientations having upper bounds lower than a previously computed estimate. This is illustrated in Figures 14-15.